EREDETI KÖZLEMÉNY

ORIGINAL ARTICLE

Interrater reliability of modified visual MRI rating scale assessing atrophy and white matter changes

Melek KANDEMIR YILMAZ[1], Zehra Betul YALCINER[2], M. Savaş TEPE[3]

[1]Bodrum American Hospital, Department of Neurology, Istanbul, Turkey
[2]Bayindir Icerenkoy Hospital, Department of Neurology, Istanbul, Turkey
[3]Bayindir Icerenkoy Hospital, Department of Radiology, Istanbul, Turkey

PRINT | English | https://doi.org/10.18071/isz.76.0019 | www.elitmed.hu

Correspondent:
Melek KANDEMIR YILMAZ, MD, MSc, Assoc. Prof., Bodrum American Hospital, Department of Neurology, Turkkuyusu Mah. Marsmabedi Cad. No: 33/35, 48400 Bodrum, Mugla, Turkey. E-mail: melekkandemir@gmail.com
https://www.orcid.org/0000-0001-7751-2002

Érkezett:
2021. szeptember 19.
Elfogadva:
2021. december 27.

Background and purpose – Cortical atrophy and white matter changes are common findings on magnetic resonance imaging among elderly. Several visual scales have been proposed to evaluate these changes using neuroimaging. We have recently proposed a scale (Modified Visual Magnetic Resonance Rating Scale) recently which allows us to evaluate atrophy, white matter hyperintensities, basal ganglia and infratentorial infarcts together. Our aim in this study was to evaluate the interrater reliability of magnetic resonance visual assessment using this scale between two neurologists and a radiologist.
Methods – Randomly selected 30 patients in different ages who underwent brain magnetic resonance imaging between January 2014 and March 2015 were included. Axial T1, coronal T2, and axial FLAIR sequences were visually scored by two neurologists and one radiologist separately. Sulcal, ventricular and medial temporal lobe atrophy, periventricular and subcortical white matter hyperintensities, basal ganglia and infratentorial infarcts were graded according to our scale. The interrater reliability and internal consistency analysis were evaluated by using intraclass correlation coefficient and Cronbach's alpha tests.
Results – The interrater agreements vary between good to excellent. The interrater correlations are moderate to excellent. Interrater correlations were excellent between two neurologists, especially on ventricular atrophy, medial temporal

Az atrófiát és a fehérállomány-változásokat értékelő módosított vizuális MRI-értékelő skála interrater megbízhatósága

Kandemir Yilmaz M,, MD, MSc; Yalciner ZB, MD; Tepe MS, MD

Háttér és cél – Kérgi atrófia és fehérállomány-elváltozások idősek esetén gyakran találhatók a mágneses rezonanciás képeken. Számos vizuális skálát javasoltak e változások idegi képalkotás segítségével történő értékelésére. Nemrégiben javasoltunk egy olyan skálát (Modified Visual Magnetic Resonance Rating Scale), ami lehetővé teszi az atrófia, a fehérállományi hiperintenzitások, a basalis ganglionok és az infratentorialis infarktusok együttes értékelését. Ebben a vizsgálatban az volt a célunk, hogy e skála segítségével két neurológus és egy radiológus bevonásával értékeljük a mágneses rezonanciás képalkotás vizuális értékelésének interrater megbízhatóságát.
Módszerek – Véletlenszerűen kiválasztott 30 különböző korú beteget vontunk be, akik 2014 januárja és 2015 márciusa között agyi mágneses rezonanciás képalkotó vizsgálaton estek át. Az axiális T1-, a coronalis T2- és az axiális FLAIR-szekvenciákat két neurológus és egy radiológus külön-külön vizuálisan értékelte, pontozta. A sulcalis, a kamrai és a medialis temporalis lebenyi atrófiát, a periventricularis és a subcorticalis fehérállományi hiperintenzitásokat, a ba-salis ganglionok és az infratentorialis rész infarktusait a mi skálánk szerint osztályozták. Az interrater megbízhatóságot és a belső konzisztencia elemzését az intraclass korrelációs együttható és Cronbach-α-tesztek segítségével értékeltük.
Eredmények – Az értékelők közötti egyezés a jó és a kiváló között változik. Az interrater korrelációk a közepestől a kiválóig terjednek. A két neurológus között kiválóak

atrophy, basal ganglia infarcts, infratentorial infarcts. When assessing ventricular atrophy, interrater correlations between individual raters were higher than sulcal atrophy. We found good correlations between neurologists and radiologist, and excellent correlations between the two neurologists for medial temporal atrophy. We found excellent interrater correlations between neurologists and radiologist for white matter hyperintensities.

Conclusion – Our scale is a reliable tool assessing both atrophy and white matter hyperintensities with a good interrater reliability. Ventricular atrophy seems to be a more reliable marker than sulcal atrophy when assessing the atrophy on neuroimaging of a patient with memory decline. We think that the total score of the scale will also guide us in clinical practice.

Keywords: cerebral atrophy, white matter hyperintensities, magnetic resonance imaging, visual grading scale, interrater reliability

voltak az interrater korrelációk, különösen a kamrai atrófia, a medialis temporalis atrófia, a basalisganglion-infarktusok és az infratentorialis infarktusok esetében. Kamrai atrófia értékelésekor az egyes értékelők közötti interrater korrelációk erősebbek voltak, mint sulcalis atrófia értékelésekor. Jó korrelációt találtunk a neurológusok és a radiológus között, és kiváló korrelációt a két neurológus között medialis temporalis atrófia esetében. Kiváló interrater korrelációt találtunk a neurológusok és a radiológus között a fehérállományi hiperintenzitások esetében.

Következtetés – Skálánk megbízható eszköz mind az atrófia, mind a fehérállományi hiperintenzitások értékelésére, jó interrater megbízhatósággal. Úgy tűnik, hogy a kamrai atrófia megbízhatóbb marker, mint a sulcalis atrófia, amikor a memóriacsökkenésben szenvedő betegeknél atrófiát értékelünk idegi képalkotás segítségével. Úgy gondoljuk, hogy a skála összpontszáma a klinikai gyakorlatban is iránymutató lesz.

Kulcsszavak: agyi atrófia, fehérállományi hiperintenzitások, mágneses rezonanciás képalkotás, vizuális osztályozási skála, interrater megbízhatóság

Sulcal atrophy (SA), ventricular enlargement, and white matter hyperintensities (WMH) are common findings of magnetic resonance imaging (MRI) among not only the elderly with dementia but also among healthy subjects[1, 2]. MRI, fluid-attenuated inversion recovery (FLAIR) sequence, in particular, is more sensitive in detecting WMH[3, 4]. We know that medial temporal lobe atrophy (MTA) is a marker for Alzheimer's Disease (AD)[5]. In addition, atrophy and WMH burden are both associated with cognitive decline[6, 7] and WMH also causes late-life depression[8, 9]. Increased WMH burden disrupts the connecting fibers causing network dysfunction[10]. Vascular risk factors, especially hypertension, and atrial fibrillation are important risk factors for atrophy and WMH[4, 11].

Various scales of similar characteristics have been developed to evaluate atrophic changes and WMH for clinical use[1, 4, 12]. WMH, observed as focal or diffuse hyperintensities on T2 and FLAIR images, is often divided into two categories: periventricular WMH (PWMH) and subcortical WMH (SCWMH)[13]. Enlargement of the sulci and ventricles is used to evaluate atrophy[14]. Basal ganglia and infratentorial hyperintensities are also assessed with visual scales[3, 4]. *Fazekas* scale is used mainly for WMH[4, 15]. Visual grading scales are based on the determination of changes in the cerebral tissue, and the atrophy and WMH[1, 4, 12, 13]. We have developed a scale called "Modified Visual MR Rating Scale (MVMRS)" which allows us to evaluate atrophy, WMH, basal ganglia, and infratentorial infarcts altogether[16].

This study aims to evaluate the interrater reliability of MRI visual assessment between two neurologists and a radiologist using the MVMRS. The work was carried out at Bayindir Icerenkoy Hospital, Department of Neurology, 34752 Atasehir/Istanbul.

Materials and methods

Randomly selected 30 patients of different ages who had brain MRI scans between January 2014 and March 2015 were included in this study. MRI scans were selected from the radiology department database. Those with extensive lesions suggestive of the cerebral vascular event (ischemic or hemorrhagic), space-occupying lesions such as tumors, lesions suggestive of demyelinating disease, or encephalitis were excluded. No clinical background was

Table 1. *Details of the MVMRS*[a]

Atrophy	Sulcal	0–9	
	Ventricular	0–9	
MTA[b]	Width of choroid fissure	Width of temporal horn	Height of hippocampal formation
0	N	N	N
1	↑	N	N
2	↑↑	↑	↓
3	↑↑↑	↑↑	↓↓
4	↑↑↑	↑↑↑	↓↓↓
WMH[c]	Periventricular	Subcortical	
	0 – No lesion 1 – Caps 2 – Thin line 3 – Halo 4– Irregular, extending to the deep white matter	0 – No lesion 1 – < 5 small focal and/or <2 large focal lesions 2 – 5–12 small focal and/or 2–4 large focal lesions 3 – >12 small focal and/or >4 large focal or confluent lesions 4 – predominantly confluent lesions	
Infarcts	Basal ganglia	Infratentorial	
	0 – No lesion 1 – Few lesions (1–3) 2 – Many lesions (>4)	0 – No lesion 1 – Few lesions (1–3) 2 – Many lesions (>4)	
Other	Tumor etc.		

[a]MVMRS: modified visual magnetic resonance imaging rating scale, [b]MTA: medial temporal lobe atrophy, [c]WMH: white matter hyperintensities

given to the evaluators so that it would not influence their decisions when evaluating the MRI. Only the MRI scans with good image quality, including axial T1 (TR= 673 ms, TE= 12 ms, 3 mm or 5 mm slice thickness, 0.6 mm gap), coronal T2 (TR= 6000 ms, TE= 90 ms, 3 mm or 5 mm slice thickness, 0.9 mm gap) and axial FLAIR (TR= 8000 ms, TE= 80 ms, 3 mm or 5 mm slice thickness, 0.6 mm gap) sequences were evaluated. All MRI scans were performed with 1.5T Siemens Magnetom Avanto (Germany). MRIs were visually scored separately by two neurologists and one radiologist using the MVMRS.

SA and ventricular atrophy (VA) were graded in a 0–9 range on the axial T1 sequences, MTA was graded in a 0–4 range on the coronal T2 sequences, PWMH and SCWMH were graded in a 0–4 range on FLAIR sequences, basal ganglia infarcts (BGI) and infratentorial infarcts (ITI) were graded in a 0–2 range on FLAIR sequences. Details about the MVMRS can be seen in our previous article[16] and are given in **Table 1**.

The study was approved by the scientific, medical ethics, and deontology board of Bayindir Hospital Icerenkoy and informed consent of all participants or their relatives were received.

The statistical analyses were performed using the Statistical Package for Social Sciences for Windows 23.0. Interrater reliability and internal consistency analyses were performed by using intraclass correlation coefficient (ICC) and Cronbach's alpha tests. In order to investigate the accuracy and validity of the parameters to be measured as well as to understand the consistency, reliability, and repeatability of the proposed scale scores, bio-statistical approaches (specifically ICC) were applied to the measurements. The Cronbach's alpha scores obtained via the reliability and validity analyses indicated the success and internal consistency of the criteria present in the scale. ICCs were designated as ≤0.40 poor to a fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement, and 0.81–1.00 excellent agreement. Cronbach's alpha designated as <0.5 unacceptable, 0.5–0.59 poor, 0.6–0.69 questionable, 0.7–0.79 acceptable, 0.8–0.89 good, ≥0.9 excellent.

Results

The mean age of the patients was 67.13 (27-91 years). There were 21 female (70%) and 9 male (30%) patients.

Table 2. *Mean and standard deviation scores of each rater for MVMRS[a] along with Cronbach's α values across raters.*

Parameters of the MVMRS	Rater 1 (Neurologist) Mean±SD	Rater 2 (Neurologist) Mean±SD	Rater 3 (Radiologist) Mean±SD	Cronbach's α values
SA[b]	4.4±1.69	4.8±2.06	4.37±2.20	0.863
VA[c]	3.17±1.76	2.87±1.98	3.93±2.21	0.975
MTA[d]	2±0.95	1.77±1.41	1.27±1.36	0.900
PWMH[e]	0.73±1.02	0.63±1.13	0.8±0.89	0.940
SCWMH[f]	1.37±1.27	1.23±1.22	1.3±0.99	0.935
BGI[g]	0.03±0.18	0.03±0.18	0.07±0.25	0.898
ITI[h]	0.2±0.48	0.13±0.35	0.17±0.46	0.933
Total score	11.90±5.38	11.4±6.31	11.90±6.86	0.971

[a]MVMRS: modified visual magnetic resonance imaging rating scale, [b]SA: sulcal atrophy, [c]VA: ventricular atrophy, [d]MTA: medial temporal lobe atrophy, [e]PWMH: periventricular white matter hyperintensities, [f]SCWMH: subcortical white matter hyperintensities, [g]BGI: basal ganglia infarcts, [h]ITI: infratentorial infarcts

The mean scores of the parameters of the MVMRS per rater are given in **Table 2**. In addition, standard deviations of the parameters were consistent across the raters. The Cronbach's α values reflecting the interrater agreement on the parameters of MVMRS varied from good to excellent. The interrater correlations, provided in **Table 3**, were generally good to excellent for the parameters of MVMRS with the exception of 0.59 SA correlation between Rater 1 and Rater 2, which may be due to the different approaches of the two specialties. According to these results, the ICC values reflect interrater reliability and internal consistency. Neuroimaging examples of SA, VA, MTA, PWMHs, and SCWMHs can be viewed in **Figures 1. A, B** and **Figures 2. A, B**.

Table 3. *Interrater correlations of MVMRS[a]*

Parameters of the MVMRS	Rater 1 vs Rater 2	Rater 1 vs Rater 3	Rater 2 vs Rater 3
SA[b]	0.588	0.643	0.807
VA[c]	0.956	0.931	0.936
MTA[d]	0.854	0.695	0.789
PWMH[e]	0.905	0.820	0.820
SCWMH[f]	0.874	0.815	0.825
BGI[g]	1	0.695	0.695
ITI[h]	0.865	0.927	0.721
Total score	0.924	0.916	0.949

[a]MVMRS: modified visual magnetic resonance imaging rating scale, [b]SA: sulcal atrophy, [c]VA: ventricular atrophy, [d]MTA: medial temporal lobe atrophy, [e]PWMH: periventricular white matter hyperintensities, [f]SCWMH: subcortical white matter hyperintensities, [g]BGI: basal ganglia infarcts, [h]ITI: infratentorial infarcts

Patients included in the study possessed small or no lesions in the BG and ITI regions in general. For these parameters, our proposed scale applies 3-class stratification in the range of 0-2. Accordingly, the mean values obtained for each rater were consistently measured as near-zero.

Discussion

It is a very challenging situation when a patient is admitted to an outpatient clinic with complaints of memory decline, which is quite common especially among the elderly. During the assessment of these patients, in addition to the clinical findings and neuropsychological tests, MRI findings can be used for guidance in determining whether the condition is degenerative or not. We have developed a practical visual MR rating scale – MVMRS – to evaluate atrophy, white matter changes, and infarcts, and found that our scale is a helpful tool in assessing the degenerative process. In this study, we indicated that the MVMRS is a reliable scale with good to excellent interrater agreements. The results of ICC and Cronbach's α tests were found to be consistent.

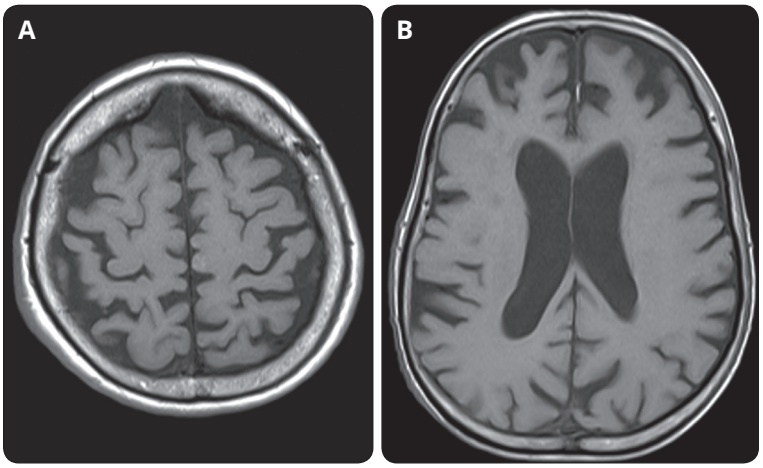

Figure 1. A. *Axial T1 image of sample with the SA score 7.* **B.** *Axial T1 image of sample with the VA score 6*

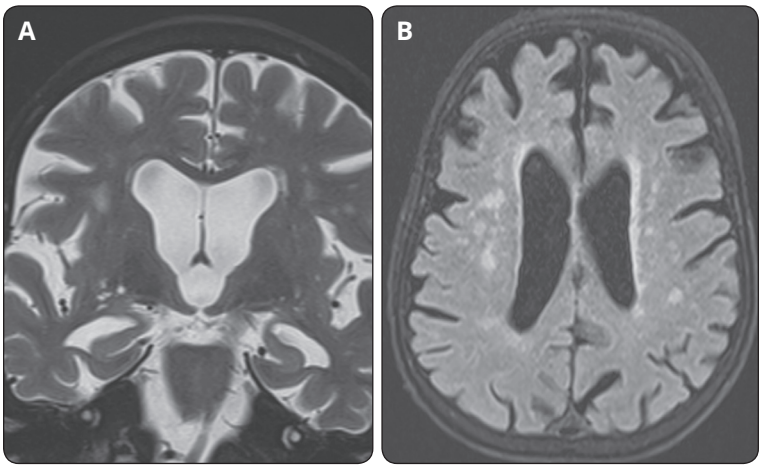

Figure 2. A. *Coronal T2 image of sample with the MTA score 3.* **B.** *FLAIR image of sample with the SCWMH score 3 and PVWMH score 2*

The scale was evaluated by two neurologists and a radiologist. Interrater correlations were excellent between the two neurologists, especially on VA, MTA, PWMH, SCWMH, BGI, ITI. When assessing VA, interrater correlations between individual raters were higher than assessing SA, and SA correlations were the lowest not just between the two neurologists, but also between each neurologist and the radiologist. According to these results, we think that the evaluation of VA is more reliable in assessing the atrophy because of the consensus between the evaluators; furthermore, it also has high sensitivity in detecting the neurodegenerative process. The slight difference between the neurologists and the radiologist in the assessment of cerebral atrophy was most likely because of the following: when evaluating and scoring cerebral atrophy, the radiologist focused on the cerebral sulci prominence, whereas the neurologists were particularly concerned about the ventricular enlargement.

When evaluating the MTA, which has high sensitivity and specificity in the identification of degenerative processes, we found good correlations between the neurologists and the radiologist, and excellent correlations between the two neurologists[16]. *Koikkalainen* et al. also reported good correlation coefficients between visual and computed rating scales for global cortical atrophy, but not as good as for MTA[17].

Most of the proposed scales make a distinction between PWMH and SCWMH without a clear definition for "small" areas and some scales ignore the periventricular thin lining[4]. It may cause some discrepancies between raters. To increase the inter- and intra-rater agreement, periventricular and deep white matter changes are assessed separately and calculated as numbers[4, 13]. Nevertheless, they found poor to reasonable inter- and intra-rater agreements[13]. We tried to define them with numbers to be more specific. We have shown high sensitivity for PWMH and SCWMH with our scale and found excellent interrater correlations between the neurologists and the radiologist[16]. We also tried to be more specific in the evaluation of the BGI and ITI, defining a few or many with numbers. Good inter- and intra-rater agreements have been found for the BGI and ITI, which can also be explained by a few infarcts in the areas of interest[13]. We found that interrater correlation between the neurologists was excellent, while it was only moderate between the neurologists and the radiologist for basal ganglia infarcts and infratentorial infarcts. This is probably because of the small sample size and the challenge in detecting small hyperintensities and differentiating them from artifacts, calcifications, or vessels in the area of interest.

The total score indicates an excellent interrater correlation between individual raters as well, pointing out a high sensitivity of the scale in assessing the pathological process[16].

While the small sample size is one of our limitations, the age range of our sample is wide. The wide age range allowed us to evaluate atrophy and WMHs of patients in various age groups. Those with extensive lesions suggestive of cerebral vascular event, space-occupying lesions, lesions suggestive of demyelinating disease, or

encephalitis were not included in this study. Therefore, there is a need for further research to evaluate their effects.

Conclusion

It is important to evaluate MRI findings with a practical tool in daily clinical practice. Our scale seems to be sensitive enough to detect WMH and atrophy with good interrater reliability in all parameters. This study shows an excellent correlation in VA, WMH, and total scores. Besides, the MVMRS provides us with a reliable evaluation of atrophy and WMH, and we think that the total score of MVMRS will also be used as guide in our clinical practice.

ACKNOWLEDGEMENT – The authors thank *Ahmet Dirican* for the support provided in the statistical analyses.
CONFLICT OF INTEREST – None.
FINANCIAL DISCLOSURE – None.

References

1. *Yue NC, Arnold AM, Longstreth WT, Elster AD, Jungreis CA, Leary DHO, et al.* Sulcal, ventricular, and white matter changes at MR imaging in the aging brain: Data from the cardiovascular health study. Radiology 1997;202:33-9. https://doi.org/10.1148/radiology.202.1.8988189
2. *Debette S, Markus HS.* The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ 2010;341:c3666. https://doi.org/10.1136/bmj.c3666
3. *Wahlund LO, Barkhof F, Fazekas F, Bronge L, Augustin M, Sjögren M, et al*. A new rating scale for age-related white matter changesapplicable to MRI and CT. Stroke 2001;32:1318-22. https://doi.org/ 10.1161/01.str.32.6.1318
4. *Mäntylä R, Erkinjuntti T, Salonen O, Aronen HJ, Peltonen T, Pohjasvaara T, et al.*Variable agreement between visual rating scales for white matter hyperintensities on MRI. Comparison of 13 rating scales in a poststroke cohort. Stroke 1997;28:1614-23. https://doi.org/10.1161/01.str.28.8.1614
5. *Visser PJ, Verhey FRJ, Hofman PAM, Scheltens P, Jolles J.* Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. J Neurol Neurosurg Psychiatry 2002;72:491-7. https://doi.org/10.1136/jnnp.72.4.491
6. *van der Flier WM, Barkhof F, Scheltens P.* Shifting paradigms in dementia. Toward stratification of diagnosis and treatment using MRI. Ann N Y Acad Sci 2007;1097:215-24. https://doi.org/10.1196/annals.1379.013
7. *Boyle PA, Yu L, Fleischman DA, Leurgans S, Yang J, Wilson RS, et al.* White matter hyperintensities, incident mild cognitive impairment, and cognitive decline in old age. Ann Clin Transl Neurol 2016;3:791-800. https://doi.org/10.1002/acn3.343
8. *Wardlaw JM, Valdés Hernández MC, Muñoz-Maniega S.* What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. J Am Heart Assoc 2015;4:001140. https://doi.org/10.1161/JAHA.114.001140
9. *Silbert LC, Nelson C, Howieson DB, Moore MM, Kaye JA*. Impact of white matter hyperintensity volume progression on rate of cognitive and motor decline. Neurology 2008;71:108-13. https://doi.org/ 10.1212/01.wnl.0000316799.86917.37
10. *Taylor ANW, Kambeitz-Ilankovic L, Gesierich B, Simon-Vermot L, Franzmeier N, Caballero MAA, et al.* Tract-specific white matter hyperintensities disrupt neural network function in Alzheimer's disease. Alzheimers Dement 2017;13:225-35. https://doi.org/10.1016/j.jalz.2016.06.2358
11. *Kandemir Yilmaz M, Yalciner ZB, Cinar S, Unay D.* Evaluating the vascular risk factors using modified visual MRI rating scale in patients with memory decline. *Under-evaluation at Clinical Neuroscience (Ideggyógyászati Szemle)*
12. *Scheltens P, Erkinjunti T, Leys D, Waklund LO, Inzitari D, del Ser T, et al.* White matter changes on CT and MRI: an overview of visual-rating scales. European task force on age-related white matter changes. Eur Neurol 1998;39:80-9. https://doi.org/ 10.1159/000007921
13. *Scheltens P, Barkhof F, Leys D, Pruvo JP, Nauta JJ, Vermersch P, et al.* A semiquantative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. J Neurol Sci 1993;114:7-12. https://doi.org/10.1016/0022-510x(93)90041-v
14. *Whitwell JL, Peterson RC, Negash S, Weigand SD, Kantarci K, Ivnik RJ, et al.* Patterns of atrophy differ among specific subtypes of mild cognitive impairment. Arch Neurol 2007;64:1130-8. https://doi.org/10.1001/archneur.64.8.1130
15. *Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA.* MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJR Am J Roentgenol 1987;149:351-6. https://doi.org/10.2214/ajr.149.2.351
16. *Yalciner BZ, Kandemir M, Taskale S, Tepe SM, Unay D.* Modified visual magnetic resonance rating scale for evaluation of patients with forgetfulness. Can J Neurol Sci 2019;46:71-8. https://doi.org/10.1017/cjn.2018.333
17. *Koikkalainen JR, Rhodius-Meester HFM, Frederiksen KS, et al; Alzheimer's Disease Neuroimaging Initiative.* Automatically computed rating scales from MRI for patients with cognitive disorders. Eur Radiol 2019;29:4937-47. https://doi.org/10.1007/s00330-019-06067-1